

Remarks

Claims 21 and 25-57 remain in the application. Claims 21 has been amended and new claims 25-57 have been added.

Applicants appreciate the courtesies extended by the Examiner and the Examiner's Primary, Mr. Allan Hoosain, to the undersigned attorney and two of the applicants, Messrs. Merrow and Kroeker, at the personal interview held at the U.S. Patent and Trademark Office on February 14, 2005. At the interview claim 21 and 25-33 were discussed. However, the Examiners informed the undersigned attorney and the attending applicants, that the prior art cited and applied in the last Official action does not teach the amended claims filed on July 19, 2004. Accordingly, all of the claims 1-24 are being resubmitted as new claims 34-57, and amended to overcome the objections raised by the for the Examiner in the outstanding Official action. Thus, all of the claims as now presented are believed allowable. If it will advance the prosecution of the application, and the Examiner believes that claims 21 and 25-33 as now submitted are allowable, Applicants are willing to cancel claims 34-57, and make them the subject of a continuation application. However, it is applicants position that all of the claims as now presented are allowable and should be allowed in the present application.

The Examiner has objected to replacement paragraphs and the amended claims 1-24 submitted in the response filed on July 19, 2004 as not being "a marked-up version to show the changes." Accordingly, the modified replacement paragraphs are being resubmitted as marked-up versions. Claims 1-24 are being resubmitted as new claims 34-57, further amended as noted below. Reconsideration of these objections is respectfully requested.

Referring to the Official action, claims 1-4, 6-14 and 16-21 have been rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement. The Examiner objected to the phrase "semantic meaning of said spoken response." While the Applicants respectfully disagree with the Examiner's rejection, in the interest of advancing prosecution, applicants have cancelled this phrase from all of the claims 34-57.

With respect to the prior art, the Examiner has entered into the record the following rejections: Claims 1, 7, 11, 17 and 21 have been rejected as being anticipated by Ljungqvist et al. Claims 2-6 and 12-16 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Ljungqvist in view of Bartholomew et al. Claims 8, 10, 18 and 20 have been rejected under

35 U.S.C. § 103(a) as being unpatentable over Ljungqvist in view of Miner et al. Claims 9 and 19 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Ljungqvist in view of Szlam et al. Claim 22 has been rejected under 35 U.S.C. § 103(a) as being unpatentable over Ljungqvist et al. in view of Bartholomew et al. Claims 23 and 24 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Matthews et al. in view of Bartholomew et al., and further in view of Brown et al. and Szlam et al. These rejections are respectfully traversed and reconsideration is requested in view of the foregoing amendments and following remarks.

Applicants' claims are directed to an outbound calling system using speaker-independent speech recognition, state-dependent responses to automatically branch the call as it progresses. More specifically, a method/system is provided for directing and controlling an outbound call, using speaker-independent speech recognition (SR) to identify the meaning of what the answering party has said. So, for example, a call can be made from a list of target people. The person is selected from the list, and a call is placed. The system can identify itself by "Hello" (or "Doe Residence", or a whole variety of other greetings) when some one picks up. The person picking up the phone may not be the target person. The system can then play "Hello, this is Eliza calling for John Doe - is he available?" The answering party can then respond with any one of an unlimited number of potential responses (such as Yes, that's me, Yes I am, Sure go ahead, No he's not, Can you hold on a minute, Can I take a message....etc.). It is not necessary to have any experience or "templates" with the answering party's voice, and the interaction (via logical branching) of the pre-recorded prompts and flexible speech recognition creates a simulated conversation between the person and the computer. The variety and combinations/complexity of the responses the system can handle is huge.

Referring to the primary reference to Ljungqvist et al., Ljungqvist et al. describe a device for conducting automatic (outbound) interactions over a telephone network. The telecommunications network/system or service described by Ljungqvist et al. is designed to deliver marketing surveys, opinion polls, advertisements, and other communications to certain targeted groups/subscribers. A corporate customer can log in to the network (via a computer interface or phone), define and order the type of interaction they want to place, and select the target group to call. Ljungqvist et al. also describe the step of determining the list of subscribers with which the interaction is to be performed via a "control node in the telecommunications network." The method/system involves logging on to the network, determining the list of people

to be called (e.g., via demographics) via this control node, calling these targets a predetermined number of times by means of the control node directing a telephone switching node, establishing a communication line with these targets if/when they pick up the phone via this switching node, determining if the interaction was "completed" with each target, and reporting the results to the customer. Also described in the reference are different types of interactions a customer might try, how to determine when or how often to call, offering gratuities/rewards to people who complete the calls, etc. The disclosure of Ljungqvist et al. does not talk at all about how the system gets the target person on the phone, if/how it asks if it has the right person, how the interaction works – none of the state modeling and descriptions that are described and claimed in the present application. Ljungqvist et al. just describes calling the targets and conducting interactions and "recording" the responses from the targets (with one of the methods for recording being speech recognition, but without describing what or how you would be recognizing during the call).

Thus, a key difference between the claimed invention and Ljungqvist et al is that the presently claimed system/method identifies the states/conditions that the system needs to manage. A model is provided for the initial portion of any outbound phone-based interaction using speech recognition. All of the different types of responses that the preferred system needs to recognize and the branching needed to actually conduct the initial portion of the interaction (asking for the correct person, recognizing a hold request, responding to "Who is calling?", leaving a message with another human if the person is not available, and a method for detecting answering machines) is disclosed. Ljungqvist et al only suggests the selection of the targets to call, and the interactions happen, without an explanation of those interactions.

It is submitted that the secondary references do not overcome the deficiencies of the Ljungqvist et al. reference.

Bartholomew et al describes a system for providing personalized calling services. The services enable a user to load/change service features into a phone switch. For example, a user could direct that their calls made from a hotel be billed to their business number, roommates on the same number could have different ring tones and billing, parents could require callers to verify their identity when calling into their home when kids are home alone (or require each kid to verify identity and then control who they can call accordingly), or to identify/screen harassing calls. Users access the system via voice verification (VV). Non-users calling a personalized line

are identified using VV as well. There are many references to voice templates and "speech based identification" - i.e., VV. The reference specifically requires that callers be instructed specifically on what to say (e.g., "This is Jane") - this is so the VV will work. There is only one potential reference to speech recognition - see column 45 lines 40-65. If VV fails on the answering party and the system hears the answering party call for the target person, it appears to use speech recognition to recognize this and then enters a hold state and waits for the VV phrase from the target ("This is Jane" again), but this is only a vague reference, and it is unclear how the speech recognition system would work. Very clearly, VV is a speaker-dependent application, as described in the Declaration under 37 C.F.R. § 1.132 by Dr. Luis Rodrigues which is being filed with this amendment. Professor Rodrigues confirms that in the industry/among those skilled in the art, the term "speech recognition" refers to speaker-independent analysis of acoustic data to determine the meaning of an utterance – and not analysis of pre-defined voice pattern templates to confirm a person's identity (this being VV).

Miner et al. describe a computer-based assistant ("Wildfire") for screening and managing inbound calls, not outbound calls. When someone calls a subscriber, the system asks the caller to state their name. It then records the name and using a speaker-dependent dictionary/contact file (i.e., names must be pre-registered) attempts to recognize the caller. If not recognized, the caller must key in their phone number. The subscriber can then choose to take the call, put it on hold, place the call into voicemail, provide for automatic call back later, etc. The reference is quite clear that voice templates/registration/training is needed for the names, and the inbound calling system is speaker-dependent..

Szlam et al. describe a method/system for telemarketers to minimize the annoyance of their hapless targets. When a telemarketer's automated dialer gets someone on the phone (no mention of speech recognition, and presumably the dialer is based on the detection of when the called party goes off hook) and an agent is not available to talk to them - it plays an apology message. There is also a mention of apologizing for wrong numbers but there is no explanation of how this type event is determined. The reference does describe using the level and duration of a signal to detect an answering machine (the present application discloses a turn taking approach). The reference also describes concealing the true purpose of the call by using children's voices in the recorded message. The Szlam et al method/system employs virtually no recognition, logic, or intelligence to function. It doesn't detect wrong numbers from telco

signals/tones or attempt to recognize when someone says so. It just plays a prompt that says "I'm sorry, I dialed the wrong number" to cover the fact that a telemarketer called and there wasn't an agent available to talk to the targeted person when the targeted person picked up the phone. The goal of doing this apparently is to prevent the targeted person from engaging call blocking, using caller ID to screen calls, #69 to identify the caller, or hanging up next time.

Matthews et al. describes a Voice Message System (VMS) for the deposit, storage, and delivery of audio (i.e., voicemail) messages. A user can deposit recorded messages in the system, and instruct it to deliver these messages to other addresses/extensions on the system. The user can also check the system for recorded messages left for him/her by other users. To gain access to the system the user must key in (touch tone) an access code and/or say a selected password (with the password being compared to an "associated distinctive voice feature template" - i.e., VV to confirm the identity of the user). When the VMS does outbound calls to deliver a recorded message, it can enable a Name Announce Feature, e.g., "This is VMS. There is a voice message for [name]." Then the user must key in his ID (touch tone) and the VMS simply plays the recorded message. There is no mention of speech recognition, or interaction/branching at all, and system is certainly not based on understanding the meaning of what the user/recipient says. The system appears to be all VV, which as described in the Rodrigues declaration is speaker-dependent.


Finally, Brown describes a message delivery system for selecting the language to be used in the system announcement before each message delivery (e.g., for international use). There is no mention of speech recognition. The only language that the Examiner references is having a user touch tone interrupt/skip as an instructional prompt. This is different from using a beep to interrupt the turn-taking approach and leaving a message on an answering machine described in the present application.

In summary, therefore, all of the claims, claims 21 and 25-57, as now presented are believed to be patentable over the cited prior art. Applicants encourage the Examiner to call the undersigned if any questions arise, or the Examiner wishes to make suggestions to advance the

prosecution of this application. Accordingly, an early and favorable action thereon, is therefore earnestly solicited. Please apply any charges or credits to deposit account 50-1133.

Respectfully submitted,

Date: 2.24.05



Toby H. Kusmer
Reg. No. 26,418
McDermott, Will & Emery
28 State Street
Boston, MA 02109
DD: (617) 535-4065
Fax: (617) 535-3800
E-mail: tkusmer@mwe.com